END
DATE
FILMED
81
DTIC

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-81-0383	2. GOVT ACCESSION NO. AD-A098076	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ROBUST ESTIMATION IN THE HETEROSCEDASTIC LINEAR MODEL WHEN THERE ARE MANY PARAMETERS		5. TYPE OF REPORT & PERIOD COVERED INTERIM rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. J. Carroll		8. CONTRACT OR GRANT NUMBER(s) AFOSR-80-0080
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of North Carolina Department of Statistics/ Chapel Hill, North Carolina		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR/NM Bolling AFB, DC 20332		12. REPORT DATE December 1980
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release;
distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Heteroscedasticity, Linear Models, Regression,
Weighted Least Squares, Robustness

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We study estimation of regression parameters in heteroscedastic linear models when the number of parameters is large. The results generalize work of Huber (1973), Yohai and Maronna (1979), and Ruppert and Carroll (1979).

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73

ROBUST ESTIMATION IN THE HETEROSCEDASTIC LINEAR MODEL WHEN THERE ARE MANY PARAMETERS

by

R.J. Carroll

University of North Carolina at Chapel Hill

Abstract

We study estimation of regression parameters in heteroscedastic linear models when the number of parameters is large. The results generalize work of Huber (1973), Yohai and Maronna (1979), and Ruppert and Carroll (1979).

Key Words and Phrases: Heteroscedasticity, Linear Models, Regression, Weighted Least Squares, Robustness.

AMS 1970 Subject Classifications: Primary 62J05; Secondary 62G35.

1. Introduction

Consider the linear model given by

$$(1.1) \qquad Y_i = \tau_i + \sigma_i \varepsilon_i \quad \text{for} \quad i = 1,\ldots,N \; ,$$

$$\tau_i = x_i \beta \; .$$

Here $\{\varepsilon_i\}$ are independently and identically distributed with symmetric distribution F, $\{x_i\}$ are $(1 \times p)$ design constants, β $(p \times 1)$ is the regression vector, and $\{\sigma_i\}$ expresses the possible heteroscedasticity in the model. Anscombe (1961) and Bickel (1978) have considered tests for the hypothesis of homoscedasticity, i.e.,

$$(1.2) \qquad H_0 \colon \sigma_1 = \sigma_2 = \ldots = \sigma_N \; .$$

In the case that (1.2) is accepted, we are in a standard linear model for which a vast theory applies. Huber (1977) and others have argued (for this homoscedastic case) that least squares methods should not be used blindly because the resulting estimates are sensitive to outliers and departures from the assumption of normal errors. Huber (1973) and many others, for example Krasker (1978), have proposed robust regression estimates. Huber (1973) and Yohai and Maronna (1979) study the asymptotic properties of these methods as $N \to \infty$ and $p \to \infty$ simultaneously. They argue that such "many parameter" asymptotics are important, both because they often occur in practice and because they show how large a sample size is needed for a given model. In effect, their work formalizes the rule of thumb: "one should take about 10 observations for every parameter."

In contrast to the homoscedastic case, there has not been much work on estimating β in the heteroscedastic case that (1.2) is rejected, and

to our knowledge there has been no work at all on the important problem of $N \to \infty$ and $p \to \infty$ simultaneously. Fuller and Rao (1978) consider the case where the Y_i occur in groups for which σ_i is constant, while Box and Hill (1974) and Jobson and Fuller (1980) assume that

$$(1.3) \qquad \sigma_i = [f(\tau_i, \theta_0)]^{-1} .$$

All three papers consider only Gaussian errors. Box and Hill suggest generalized weighted least squares (WLS) wherein the weights (1.3) are estimated and WLS applied. Jobson and Fuller consider maximum likelihood techniques which are particularly sensitive to non-normal errors as well as small misspecifications in (1.3).

For the case p fixed and $N \to \infty$, Ruppert and Carroll (1979) developed a class of regression estimates for the heteroscedastic linear model, which are both robust and do not depend on the normal error assumption. Their estimates are a natural generalization of Huber's Proposal 2 (1977) and are defined as follows. First, assume

$$(1.4) \qquad f(\tau, \theta) = \exp(\theta h(\tau)) , \theta(1 \times r) .$$

This class includes the important special cases

$$(1.5) \qquad f(\tau, 0) = \sigma(1 + \tau^2)^{\alpha/2} \quad \text{or} \quad \sigma \exp(\alpha \tau) \quad \text{or} \quad \sigma |\tau|^\alpha .$$

Let ψ be an odd function and χ be an even function $(\chi(-\infty) < 0 , \chi(+\infty) > 0)$. Let $\hat{\beta}_p$ be a preliminary estimate of β and define $\hat{\theta}$ as any solution to

$$(1.6) \qquad H_N(\theta) = N^{-1} \sum_{i=1}^{N} \chi((Y_i - t_i) f(t_i, \theta)) h(t_i) = 0 ,$$

where $t_i = x_i \hat{\beta}_p$. If an exact solution to (1.6) does not exist, define $\hat{\theta}$ to minimize $|H_N(\theta)|$. Finally, define $\hat{\beta}$ as the solution to

$$(1.7) \qquad N^{-1} \sum_{i=1}^{N} \psi((Y_i - x_i\beta)f(t_i,\hat{\theta}))x_i f(t_i,\hat{\theta}) = 0 \ .$$

We make the conventions

$$(1.8) \qquad \begin{aligned} \theta_0 &= (-\log \sigma_0 \quad \theta_{20}\ldots\theta_{r0}) \\ h(\tau) &= (1 \quad h_2(\tau)\ldots h_r(\tau)) \ , \end{aligned}$$

with the parameter σ_0 chosen so that

$$(1.9) \qquad Ex((Y_i - \tau_i)f(\tau_i,\theta_0)) = Ex(\epsilon_i) = 0 \ .$$

The conventions (1.8)-(1.9) ensure consistency of the solutions. In the case of homoscedasticity, the solution to (1.6)-(1.7) is trivially Huber's Proposal 2 if the preliminary estimate $\hat{\beta}_p$ is also. Ruppert and Carroll (1979) show that, in general, the solution to (1.6) and (1.7) is asymptotically equivalent to the "optimal" robustified WLS obtainable when the $\{\sigma_i\}$ are actually known, i.e.,

$$L(N^{\frac{1}{2}}(\hat{\beta}-\beta)) \to N_p(0, S^{-1} E\psi^2(\epsilon_1)/(E\psi'(\epsilon_1))^2) \ ,$$

$$(1.10) \qquad S = \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} x_i' x_i f^2(\tau_i,\theta) = \lim_{N\to\infty} S_N \ ,$$

$$N_p(0,V) = \text{p-variate normal, mean 0, covariance V} \ .$$

Monte-Carlo results show that the asymptotics are surprisingly accurate for small samples.

The purpose of this paper is to generalize the results of Ruppert and Carroll to the case $N \to \infty$, $p \to \infty$, under conditions similar to those needed by Bickel (1978) and Yohai and Maronna (1979). In the next section we consider the estimates of β when an appropriate estimate of θ is available, and in Section 3 we consider the estimation of θ.

2. The Limit Distribution of the Regression Estimate

Assuming that $\hat{\theta}$ is of the correct order ((A10) below) and $h(\cdot)$ is sufficiently smooth ((A3) below), one can follow the steps of Yohai and Maronna (1979). We first establish the order of $\hat{\beta}$.

Theorem 1. *Under* (A1)-(A11) *below,*

$$N|\hat{\beta} - \beta|^2/p_N = O_p(1) .$$

The main result is as follows.

Theorem 2. *Assume* (A1)-(A12) *below. Let* V_N *be any symmetric square root of* S_N *((1.10) and (A11)) and let* $\{a_N\}$ *be a sequence of* $(p_N \times 1)$ *vectors with* $|a_N| = 1$. *Then*

$$N^{\frac{1}{2}} a_N' V_N(\hat{\beta} - \beta) \xrightarrow{L} N(0 , E\psi^2(\varepsilon_1)/(E\psi'(\varepsilon_1))^2) .$$

Here are the assumptions.

(A1) $\quad N^{-1} \sum\limits_{i=1}^{N} x_i' x_i = I .$

(A2) $\quad |\tau_i| \leq M < \infty .$

(A3) The function $h(x)$ and its first two derivatives are continuous and bounded on the set $|x| \le M + \epsilon$ for some $\epsilon > 0$. Further, $M_1^{-1} < f(x,\theta) \le M_1$ on this set uniformly for $|\theta - \theta_0| \le \epsilon$.

(A4) $\lim\limits_{N \to \infty}(p_N^{3/2}/N) \max\limits_{1 \le i \le N} |x_i|^2 = 0$.

(A5) ψ is non-decreasing and bounded.

(A6) ψ is Lipschitz and constant outside an interval.

(A7) If $D(u,z) = (\psi(u+z) - \psi(u))/z$, then there exists b, c, and d such that $P(|\epsilon_1| < cM_1^{-2}) > 0$ and $D(u,z) \ge d$ if $|u| \le c$, $|z| \le b$.

(A8) ψ and its first three derivatives are continuous and bounded.

(A9) $N|\hat{\beta}_p - \beta|^2/p_N = 0_p(1)$.

(A10) $N|\hat{\theta} - \theta|^2/p_N = 0_p(1)$.

(A11) $0 < \lim\inf \lambda_{min}(S_N) \le \lim\sup \lambda_{max}(S_N) < \infty$, where $\lambda_{min}(\cdot)$ and $\lambda_{max}(\cdot)$ are the minimum and maximum eigenvalues respectively.

(A12) $T_N \overset{p}{\to} 0$ where $T_N = (\hat{\beta}_p - \beta)' W_0(\hat{\beta} - \beta)$,

$$W_0 = N^{-\frac{1}{2}} \sum_1^N f^2(\tau_i, \theta_0) a_N' V_N^{-1} x_i' x_i' x_i \frac{d}{d\tau} f(\tau_i, \theta_0).$$

Remarks on Assumptions: Assumption (A1) is a reparameterization device used by Huber (1973) and Bickel (1978). Assumption (A2), which is also used by Bickel, can be deleted if one strengthens (A3) by making $M = \infty$, a set of circumstances holding for the homoscedastic case. Assumption

(A3) is stronger than required but is quite convenient and is true for the first two special cases in (1.5). Assumptions (A4)-(A8) are as in Yohai and Maronna (1979) except for the second part of (A6), which holds for those ψ commonly used and which can be relaxed at the cost of some technical complications. Part (A8), in particular, can be relaxed if $p_N^{3/2}$ in (A4) is strengthened to p_N^4; the proof becomes quite messy. Part (A9) holds, for example, in cases of homoscedasticity or, in general, for least squares if $E\varepsilon_1^2 < \infty$, and (A10) will be verified in the next section. Part (A11), in effect, formalizes (1.8) and follows from (A1) in the important special case of homoscedasticity.

Assumption (A12) is somewhat awkward looking and is the only obstacle to our Theorem 2 being a real generalization of previous work. However, (A12) holds in important special cases, including homoscedasticity, as the following shows.

Proposition 1. *Assumption* (A12) *holds if any of the following obtain:*

(2.1)   *The variances are homoscedastic.*

(2.2) $(p_N^2/N)\max|x_i|^2 \to 0.$

(2.3) $E\varepsilon_1^2 < \infty$, $\hat{\beta}_p$ *is the least squares estimate, and*

$$\lim \sup \lambda_{max}(N^{-1} \sum_1^N x_i' x_i\, f(\tau_i,\theta_0)) < \infty.$$

Proof of Theorem 1. The proof parallels that of Theorem 2.2 of Yohai and Maronna. Let $Y_i - \tau_i = z_i = \varepsilon_i/f(\tau_i,\theta_0)$ and define

$$U_N(\gamma, L) = p_N^{-1/2} \sum_{i=1}^N \psi((z_i - Ld_i \gamma)f(t_i,\hat{\theta}))d_i \gamma f(t_i,\hat{\theta}) ,$$

$$d_i = x_i N^{-1/2} .$$

Since ψ is monotone,

$$P\{N|\hat{\beta} - \beta|^2/p_N > L^2\} \le P\{\sup U(\gamma, Lp_N^{\frac{1}{2}}) > 0, |\gamma| = 1\} \ .$$

Now

$$U(\gamma, Lp_N^{\frac{1}{2}}) = A_{N1}\gamma - A_{N2}(\gamma) \ ,$$

where

$$A_{N1}\gamma = p_N^{-\frac{1}{2}} \sum_1^N \psi(z_i f(t_i,\hat{\theta}))d_i \gamma f(t_i,\hat{\theta})$$

$$A_{N2}(\gamma) = L \sum_{i=1}^N D(z_i f(t_i,\hat{\theta}), -Ld_i \gamma f(t_i,\hat{\theta})p_N^{\frac{1}{2}})(d_i \gamma)^2 f^2(t_i,\hat{\theta}) \ .$$

From (A3) and (A4) it follows that

$$\sup\{|d_i \gamma f(t_i,\hat{\theta})p_N^{\frac{1}{2}}| : |\gamma| = 1, 1 \le i \le N\} \overset{p}{\to} 0 \ .$$

From (A3) we find that

$$A_{N2}(\gamma) \ge (L d/M_1) \sum_{i=1}^N (d_i \gamma)^2 I(|\varepsilon_i| \le cM_1^{-2})$$

$$\ge (Ld/M_1)P(|\varepsilon_1| \le cM_1^{-2}) + o_p(1) \ ,$$

the last step following from Y-M and (A7). We thus need only prove
$A_{N1}\gamma = 0_p(1)$. Rewrite $A_{N1}\gamma$ as

$$A_{N1}\gamma = B_N\gamma + (A_{N1}-B_N)\gamma \ ,$$

$$B_N = (Np_N)^{-\frac{1}{2}} \sum_{i=1}^N \psi(z_i)x_i f(\tau_i,\theta_0) \ .$$

By (A5) and (A11), $E|B_N|^2 = 0(1)$ so that

$$\sup\{|B_N\gamma|: |\gamma| = 1\} = 0_p(1) \ .$$

By (A2), (A3), (A6) and (A8), for some $M_2 > 0$,

$$|(A_{N1}-B_N)\gamma| \le M_2(Np_N)^{-\frac{1}{2}} \sum_{i=1}^{N} |x_i\gamma|(|t_i \cdot \tau_i| + |\hat{\theta}-\theta|) \ .$$

By (A1), (A9) and (A10), this is $0_p(1)$ uniformly for $|\gamma| = 1$, completing the proof. $\square$

<u>Proof of Theorem 2</u>. The proof is similar to that of Theorem 3.2 of Yohai and Maronna. By a Taylor expansion,

$$0 = \sum_1^N \psi((Y_i - x_i\hat{\beta})f(t_i,\hat{\theta}))a_N'V_N^{-1}x_i'q_i$$

$$= w_1 - w_2 - w_3'(\hat{\beta}-\beta) + \frac{1}{2}(\hat{\beta}-\beta)' W_4(\hat{\beta}-\beta) - \frac{1}{6} w_5$$

where

$$q_i = f(t_i,\hat{\theta}), \quad r_i = f(\tau_i,\theta_0)$$

$$w_1 = N^{-\frac{1}{2}} \sum_1^N \psi(z_iq_i)q_ia_N'V_N^{-1}x_i'$$

$$w_2 = (E\psi')N^{\frac{1}{2}}a_N'V_N(\hat{\beta}-\beta)$$

$$w_3' = N^{-\frac{1}{2}} \sum_1^N (\psi'(z_iq_i)q_i^2 - (E\psi')r_i^2)a_N'V_N^{-1}x_i'x_i$$

$$W_4 = N^{-\frac{1}{2}} \sum_1^N \psi''(z_iq_i)a_N'V_N^{-1}x_i'x_i'x_iq_i^3$$

$$|w_5| = 0(1)|N^{-\frac{1}{2}} \sum_1^N a_N'V_N^{-1}x_i'(x_i(\hat{\beta}-\beta))^3| \xrightarrow{p} 0 \ .$$

Note that from (A2), (A3), and (A10),

$$(2.4) \qquad |q_i - r_i| = \mathcal{O}(|t_i - \tau_i| + |\theta - \theta|) = \mathcal{O}(|t_i - \tau_i| + (p_N/N)^{\frac{1}{2}}) .$$

Then (2.4), (A8), and Theorem 1 show that

$$(2.5) \qquad (\hat{\beta}-\beta)' W_4 (\hat{\beta}-\beta) = N^{-\frac{1}{2}} \sum_1^N \psi''(\epsilon_i) r_i^2 (x_i(\hat{\beta}-\beta))^2 a_N' V_N^{-1} x_i' + R_{N1} ,$$

$$(2.6) \qquad |R_{N1}| = \mathcal{O}(1) N^{-\frac{1}{2}} \sum_1^N (x_i(\hat{\beta}-\beta))^2 |q_i - r_i| |a_N' V_N^{-1} x_i'| \xrightarrow{p} 0 .$$

Thus, as in (3.5) and (3.6) of Yohai and Maronna, we obtain

$$(2.7) \qquad (\hat{\beta}-\beta)' W_4 (\hat{\beta}-\beta) \xrightarrow{p} 0 ,$$

$$(2.8) \qquad w_3'(\hat{\beta}-\beta) = R_{N2} + o_p(1) ,$$

$$R_{N2} = N^{-\frac{1}{2}} \sum_1^N (\psi'(z_i q_i) q_i^2 - \psi'(\epsilon_i) r_i^2) a_N' V_N^{-1} x_i' x_i (\hat{\beta}-\beta) .$$

By Taylor expansions of ψ' and f, we obtain

$$w_3'(\hat{\beta}-\beta) = (2E\psi' + E\epsilon_1 \psi''(\epsilon_1)) T_N + o_p(1) ,$$

where T_N is given in (A12). Because of (A12), we see that the proof is completed by showing

$$(2.9) \qquad w_1 = N^{-\frac{1}{2}} \sum_1^N \psi(\epsilon_i) r_i a_N' V_N^{-1} x_i' + o_p(1) .$$

By using the symmetry of F, the fact that ψ is odd, and as in (3.5) and (3.6) of Yohai and Maronna, (2.9) follows. □

Proof of Proposition 1. Under (2.1),

$$\frac{d}{d\tau} f(\tau,\theta_0) = 0 ,$$

verifying (A12). A similar easy proof follows from (2.2). When (2.3) holds,

$$|T_N|^2 = O_p((p_N/N))D_N ,$$

$$D_N = \sum_1^N (a_N'V_N^{-1} x_i)^2 (x_i(\hat{\beta}_p-\beta))^2 ,$$

$$ED_N \le N^{-1} \max|x_i|^2 \sum_1^N (a_N'V_N^{-1} x_i)^2 = O(\max|x_i|^2) ,$$

completing the proof by (A4). □

3. The Order of $\hat{\theta}$

From Theorems 1 and 2, we see that we must exhibit a sequence of estimates $\hat{\theta}$ satisfying assumption (A10). We will show that any solution to (1.6) will work under the following assumptions.

(B1) The even function χ satisfies the same assumptions as does ψ.

(B2) The statistic

$$G_N = \theta_0 N^{-1} \sum_{i=1}^{N} h(\tau_i)(t_i-\tau_i) = O_p((p_N/N)^{\frac{1}{2}}) .$$

(B3) $0 < \lambda_\infty = \lim \inf \lambda_{min}(Q_N) \le \lim \sup \lambda_{max}(Q_N) < \infty ,$

$$Q_N = (E\epsilon_1 \chi'(\epsilon_1))N^{-1} \sum_1^N h(\tau_i)^T h(\tau_i) .$$

We note that, when r is fixed, (B2) holds when $\hat{\beta}_p$ is the least squares estimate, while Huber's (1973) proof can be used for classical

M-estimates if (2.2) holds. Also, (B2) follows from assumption (A9) if there is homoscedasticity; for in this case,

$$\theta_0 = (\log \sigma \quad 0 \ldots 0)$$

$$h(\tau) = (1 \quad h_2(\tau) \ldots h_r(\tau))$$

$$G_N = (\log \sigma) N^{-1} \sum_{i=1}^{N} (t_i - \tau_i) \ .$$

Theorem 3. *Assume* (A1)-(A9), (A11)-(A12), *and* (B1)-(B3). *Then* (A10) *holds.*

We need the following lemma.

Lemma 1. *Let* $\{v_i\}$ *be i.i.d. mean-zero bounded random variables; let* $\{s_i\}$ *be uniformly bounded constants. Then*

$$(3.1) \qquad A_N = N^{-1} \sum_{i=1}^{N} v_i s_i (t_i - \tau_i) = O_p(p_N/N) \ .$$

Proof. From (A9),

$$A_N^2 \le |N^{-1} \sum_{i=1}^{N} v_i s_i x_i|^2 \, O_p(p_N/N) = D_N O_p(p_N/N)$$

$$E D_N = O(1) p_N/N \ .$$

Proof of Theorem 3. Define $H_N(\cdot)$ as in (1.6) and note that by (A1), (A4), and (A9),

$$\max\{|t_i - \tau_i| : 1 \le i \le N\} \xrightarrow{p} 0$$

$$(3.2)$$

$$N^{-1} \sum_{i=1}^{N} (t_i - \tau_i)^2 = O_p(p_N/N) \ .$$

From (A4), (B1) and Lemma 1, we thus obtain

$$(3.3) \qquad H_N(\theta_0) = N^{-1} \sum_{i=1}^{N} \chi(\varepsilon_i) h(\tau_i) + G_N E \varepsilon_1 \chi'(\varepsilon_1) + 0_p(p_N/N)$$

$$= 0_p((p_N/N)^{\frac{1}{2}})$$

from (B1), (B2) and the fact that $E\chi(\varepsilon_1) = 0$. By (A4), (B1)-(B3), (3.2) and a Taylor expansion, we find that for any $M_2 > 0$,

$$(3.4) \qquad N^{\frac{1}{2}} \sup\{|H_N(\theta_0 + \Delta N^{-\frac{1}{2}}) - H_N(\theta_0) - \Delta Q_N N^{-\frac{1}{2}}| : |\Delta| \le M_2 p_N^{\frac{1}{2}}\} \xrightarrow{p} 0 \ .$$

Now fix $\varepsilon > 0$, $L > 0$. Choose γ, N_0, D_0 such that $N \ge N_0$ implies

$$P\{N^{\frac{1}{2}}|G_N| \ge D_0 p_N^{\frac{1}{2}}\} < \varepsilon/3$$

$$P\{N^{-\frac{1}{2}}|\sum_1^N \chi(\varepsilon_i) h(\tau_i)| \ge \gamma\} < \varepsilon/3 \ .$$

Define M_2 by

$$L = \lambda_\infty M_2/2 - \gamma p_N^{-\frac{1}{2}} - D_0 \ .$$

Choose $N_1 \ge N_0$ so that $N \ge N_1$ implies

$$\lambda_{\min}(Q_N) \ge \lambda_\infty/2$$

and

$$P\{\text{left side of } (3.4) \ge \gamma/2\} < \varepsilon/3 \ .$$

Then with probability at least $1-\varepsilon$, $N \ge N_1$ and $|\Delta| = M_2 p_N^{\frac{1}{2}}$ imply

$$(3.5) \qquad (N/p_N)^{\frac{1}{2}} \, \Delta H_N(\theta_0 + \Delta N^{-\frac{1}{2}}) \geq M_2 \, L p_N^{\frac{1}{2}} \; .$$

Also note that $\Delta H_N(\theta_0 + s \Delta N^{-\frac{1}{2}})$ is increasing as a function of s; so that if $|\Delta| \geq M_2 p_N^{\frac{1}{2}}$, then as in Jurecková's (1977) proof, using (3.5),

$$(3.6) \qquad (N/p_N)^{\frac{1}{2}} \, \Delta H_N(\theta_0 + \Delta N^{-\frac{1}{2}}) \geq |\Delta| L p_N^{\frac{1}{2}} \; .$$

From (3.6), we have with probability at least $1-\varepsilon$,

$$\inf\{p_N^{-\frac{1}{2}} |H_N(\theta)| : N^{\frac{1}{2}} |\theta - \theta_0| \geq M_2 p_N^{\frac{1}{2}}\}$$

$$\geq \inf\{|\Delta H_N(\theta_0 + \Delta N^{-\frac{1}{2}})| / |\Delta| p_N^{\frac{1}{2}} : |\Delta| \geq M_2 p_N^{\frac{1}{2}}\}$$

$$\geq L \; .$$

Since L is arbitrary, the proof is complete. $\qquad\qquad\square$

Remark: Throughout we have taken r fixed; r is the dimension of θ_0 and $h(\tau)$. This seems a perfectly reasonable assumption, as one is unlikely to construct a complicated function for σ_i, and in fact, $r \leq 3$ will suffice for most applications. However, the proofs can be forced through when $r = r_N = O(p_N^{\frac{1}{2}})$. One must assume that the bounds in (A3) are uniform in the components of $h(\cdot)$, and the only stumbling block is the assumption (B2). When $\hat{\beta}_p$ is the least squares estimate and $E\varepsilon_1^2 < \infty$, it turns out that (B2) holds whenever

$$(Np_N^{\frac{1}{2}})^{-1} \sum_1^N |h_i| |c_i| = O(1) \; .$$

References

Anscombe, F.J. (1961). Estimation of residuals. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* (J. Neyman, Editor), pp. 1-36. University of California Press, Berkeley.

Bickel, P.J. (1978). Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. *Ann. Statist. 6*, pp. 266-291.

Box, G.E.P. and Hill, W.J. (1974). Correcting inhomogeneity of variances with power transformation weighting. *Technometrics 16*, pp. 385-389.

Fuller, W.A. and Rao, J.N.K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Ann. Statist. 6*, pp. 1149-1158.

Huber, P.J. (1973). Robust regression: Asymptotics, conjectures and Monte-Carlo. *Ann. Statist. 1*, pp. 799-821.

Huber, P.J. (1977). *Robust Statistical Procedures*. SIAM, Philadelphia.

Jobson, J.D. and Fuller, W.A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *J. Am. Statist. Assoc. 75*, 176-181.

Jureckova, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear regression models. *Ann. Statist. 5*, 464-472.

Krasker, W.S. (1978). Estimation in linear regression models with disparate data points. To appear in *Econometrica*.

Ruppert, D. and Carroll, R.J. (1979). M-estimates for the heteroscedastic linear model. *Mimeo Series #1243*, University of North Carolina.

Yohai, V.J. and Maronna, R.A. (1979). Asymptotic behavior of M-estimates for the linear model. *Ann. Statist. 7*, 258-268.

DATE FILMED

8